Manas Journal of Engineering
ISSN 1694-7398 | e-ISSN 1694-7398

Volume 7 (Issue 2) (2019) Pages 141-146

Assisting tool for essay grading for Turkish language instructors

Mustafa Alp Çetin[1], Rita Ismailova[2,*]

[1]Kyrgyz-Turkish Manas University, Computer Engineering Department, Bishkek, Kyrgyzstan, mustafaalpcetin1@gmail.com

[2]Kyrgyz-Turkish Manas University, Computer Engineering Department, Bishkek, Kyrgyzstan, rita.ismailova@manas.edu.kg, ORCID: 0000-0003-0308-2315

ABSTRACT

When learning languages, writing an essay is one of the main methods for assessing students' knowledge. However, with the development of information and communication technologies, language learning is also being transferred to online platforms. At the same time, as the number of students increases, the problem of evaluating students' essays arises. In this paper, we offer an automated system that facilitates instructors while evaluating students' essays. Currently, the system works for essays written in Turkish. The system was built using the Zemberek library. It allows one to extract text features the essay of several people at the same time on several indicators, namely, morphological analysis, vocabulary, the use of different language structures, etc. Currently, many automated essay grading tools are proposed, and one of the main factors that defined their accuracy it the extraction of text features. Thus, as further work, it is planned to use the data obtained using this essay assessment system together with instructors' evaluation to create an expert system for automatic essay evaluation using machine learning techniques.

ARTICLE INFO

Research article

Received: 17.12.2019
Accepted: 23.12.2019

Keywords:
Natural language processing, automated tool, essay evaluation, Computer-assisted language learning

*Corresponding author

1. Introduction

Nowadays, natural language processing is becoming a widespread subject and many types of research are being carried out in this field. This technique deals with the conversion of human language into a form that eases the computer manipulations on the language [9]. Another definition by Chowdhury (2003) emphasizes that the technique of natural language processing aims at developing tools so that computer systems could manipulate natural languages to perform desired tasks [8]. Processing techniques differ by their complexity, starting from text categorization and word frequency count [5, 7, 10, 12] to text translation [2, 4, 13, 18, 23] smart text annotation [21] and generating meaningful responses to human questions [3]. As natural language processing techniques develop, its usage area is also expanding.

Despite the popularity of natural language processing, it has not been fully completed for any language today. The most NLP solutions are offered for the English language. Though there are studies presented for other languages as well [1, 11, 24].

Looking at the application of NLP from the other end, and considering the way people learn languages, essay writing is one of the most effective methods that allow evaluating how much the language learners are adapting to the language they are learning. However, evaluation and grading written works is a hard task, especially as the number of students increases. Besides, human rating in some cases can be considered as bias [25]. In addition, grading essays is an expensive task [14]. Especially this issue becomes very acute when using massive open online courses, where students expect feedback for their writing assignments.

In this article, the use of language processing techniques for the evaluation of essay assignments is considered. We aimed to develop a fast automatic evaluation system by using Turkish natural language processing tools. For that, the Zemberek Turkish natural language processing library by Akın and Akın, developed in 2007, and its modules were utilized for the morphological analysis of essays written in Turkish. However, since prose evaluation is more than a count of a spelling error, in the frame of the current study, the software was developed for segregating the text assignments.

Thus, the tool serves as an assisting tool for language teachers.

The article is organized as follows. Section 2 starts with an overview of essay grading techniques and software. After that, the methodology, utilized in the current work is described along with materials. Results are presented in Section 4.Section 5 concludes the work.

2. Related works

An essay evaluation program was first proposed in 1966 by Page [14, 15]. The author proposed to grade essays in two dimensions – by analyzing content and writing style. In order to predict an essay grading, initially, 272 essays were evaluated by four independent instructors. Next, the text features such as number of words, number of parts of speech or length of words of the essays were extracted and analyzed by a multiple regression method. Finally, a multivariate relationship between human and machine evaluations was analyses to make essay grading more accurate. With the development of computer science, this approach was further modified by the author by adding grammar checking, dictionaries, tagger and parsers [16].

As for the practical implementation of essay grading software, Shermis et al. proposed a web-system, which served for placement tests for English learners by extending the idea by Page et al. in 1997 [17]. The system was tested based on the work of 807 students' text assignments. The results of the study suggested that computer evaluation was as accurate, as that of humans [20].

In 2003, Burstein et al. [6] also proposed an online application consisting of two components, an essay scoring component, and a writing analysis tool. The scoring component was built using content vector analysis while the writing analysis part utilized NLP and statistical machine learning techniques. The evaluation was done based on the grammar, repetition of word usage and disclosure structures. As early essay evaluation was mostly based on spelling error count, Schraudner proposed to correct students' errors by collecting students' assignments using Google forms, which were further analyzed using three different online services for correcting grammar errors [19].

However, later works in this field are done using achievements in machine learning. For example, the automated essay scoring system, proposed by Taghipour and Ng in 2016 utilizes recurrent neural networks [22]. Zupanc and Bosnić in 2017 proposed a system that increases the evaluation accuracy by enhancing the semantic scoring [26]. Thus, as can be seen, mostly algorithms that serve as a basis for written task evaluation are based on partitioning and putting prose into some framework [20]. Thus, the correct partitioning can serve as a good basis for the further text assignment evaluation system.

3. Materials and methodology

3.1. Materials and Workflow

The system has been developed in NetBeans development environment using Java programming language. In addition, for the language processing, Morphology, Tokenization and Normalization modules of Zemberek Turkish natural language processing framework by Akın and Akın, developed in 2007 [1] was utilized. The choice of these modules was due to a task that was aimed while developing the software.

For the evaluation, first, the text is divided into sentences by the Tokenization module and the incorrect usage within the sentence is corrected with the Normalization module. However, before normalization, these usages are counted, and the number of errors is displayed in an output file. Then, all the elements of the sentence are separated using the Morphology module; the following morphological outputs are determined:

- the morphological output of the vocabulary used in the text;
- morphological word wealth (parts of spech);
- use of indicative moods (tenses);
- use of subjunctive moods;
- the total number of sentences used;
- the total number of words used.

The choice of outputs was due to the scaling used by the language preparatory school. However, since some evaluation scales such as relevance or redundancy of words and sentences or disclosure of the topic of the essay could not be implemented, the list of outputs was limited to the above measurements. Yet, the keywords entered by the instructor can be identified in the list of vocabulary knowledge obtained as a result of these processes.

In addition, to detect incorrect spelling of words in the text the Turkish Spell Checker method within the normalization module was utilized. The data obtained here is saved in the Excel file format using the Apache POI application programming interface.

3.2. Research Hypotheses and Questions

As was mentioned above, the project aimed to develop an essay evaluation software that could assist language instructors in the evaluation of essay assignments. However, to be able to build a system that would have a high accuracy in the essay grading, the proper component extraction is needed. Therefore, in the scope of current work, the text

feature extraction was carried out. While designing the tool, the following research questions were formulated:

1. What are the specific criteria for essay evaluation?
2. What methods of natural language processing can be used for essay evaluation?
3. What barriers does one face while using Turkish natural language processing methods?

In the current study, the first research question was considered. However, results obtained by answering the first research question would shed a light to the hypotheses for further works formulated as follows:

1. By evaluating the essay assignment of the language learners, an instructor can comment on how much the student mastered the subject based on the use (or disuse) of keywords, provided by the instructor.
2. By evaluating the essay assignment of the language learners, an instructor can determine exactly where the student has deficiencies in language learning through the spelling errors that the student has made repeatedly.
3. By evaluating the essay assignment of the language learners, an instructor can measure by how many different words the student can express himself/herself and measure the sufficiency of the vocabulary wealth on the given subject. In addition, these words can be analyzed morphologically by counting the number of nouns, verbs, adjectives and so on. Thus, the morphological correctness of sentences can be observed.
4. Since in the Turkish language, multiple, recursively addable derivational suffixes are used for word formation, by determining the modalities of indicative moods and subjunctive moods usage in the Turkish language, an instructor can measure how many moods the student can use while writing an essay in Turkish and the ability to use certain Turkish suffixes.
5. Comparing the total number of unique words that a student used in an essay with the total number of words used, an instructor can measure whether the student constantly makes statements using the same words.

The system, developed in the frame of the current study, provides a technical background to estimating these hypotheses.

4. Results

The software, developed for essay feature extraction, works on a base of written assignments. The developed tool takes as input files in XML format. Thus, for essay submission, the special text editor was developed, where essays are recorded in XML file format. In addition, the system allows the instructor (or system administrator) to include several essays (files) and keywords in the main program as well (Fig. 1).

The results of the essay evaluation are recorded in the Excel file format. The system separates each submission by recording them in separate rows, thus, several files can be graded simultaneously. However, although the feature extraction is done for each file (essay) separately, an output is provided in a single file, with each essay features in one row. In addition, it is possible to see the program output on the text editor screen as well as demonstrated in Fig. 1

In the Excel file, there are several fields, reflecting identification information such as the student's surname, student ID number; the next sections include information related to the current essay topic in form of several keywords. While extracting features, the program counts keywords' usage by language learners and provides this information to the instructor. Thus, the instructor can check the relevance of the essay to the given topic.

Grammar features of text assignments are provided in the field reflecting spelling errors. Finally, the semantic structure of sentences, used by language learners are given by morphological vocabularies, the number of sentences using indicative moods and subjunctive moods, the total number of words, the total number of sentences and evaluation information, which can be seen in separate rows for each student (Table 1).

.

Table 1. *Sample software output in excel format (the simplified transposed form)*

Name Surname	Student 1	Student 2	Student 3	Student 4
Student ID	xxxx.xxxx	xxxx.xxxx	xxxx.xxxx	xxxx.xxxx
Keywords	gelmek gitmek dil öğrenmek	gelmek gitmek dil öğrenmek	gelmek gitmek dil öğrenmek	gelmek gitmek dil öğrenmek
Spelling errors	uyanırır, baktığınğz, bakasanız, … Number of errors: 10	şoyle, büyüklerimzden, soyler, … Number of errors: 14	Hintler, Çinliler, Amerikalılarla , … Number of errors: 21	Bügün, dı, hokabaz, sehitliğe , … Number of errors: 9
Morphological Word Wealth	Nouns: 45 Verbs: 33 … Unknown: 1	Nouns: 31 Verbs: 24 … Unknown : 0	Nouns: 82 Verbs: 49 … Unknown: 1	Nouns: 79 Verbs: 63 … Unknown: 1
Vocabulary Wealth	// unique words listed // number of unique words			
indicative moods (tenses)	// list of moods // the number of sentences in an indicative moods			
Subjunctive moods	// list of moods // the number of sentences in a subjunctive mood			
Total Words Used	176	140	432	465
Total Number of Sentences	27	16	67	75
Evaluation	-	-	-	-

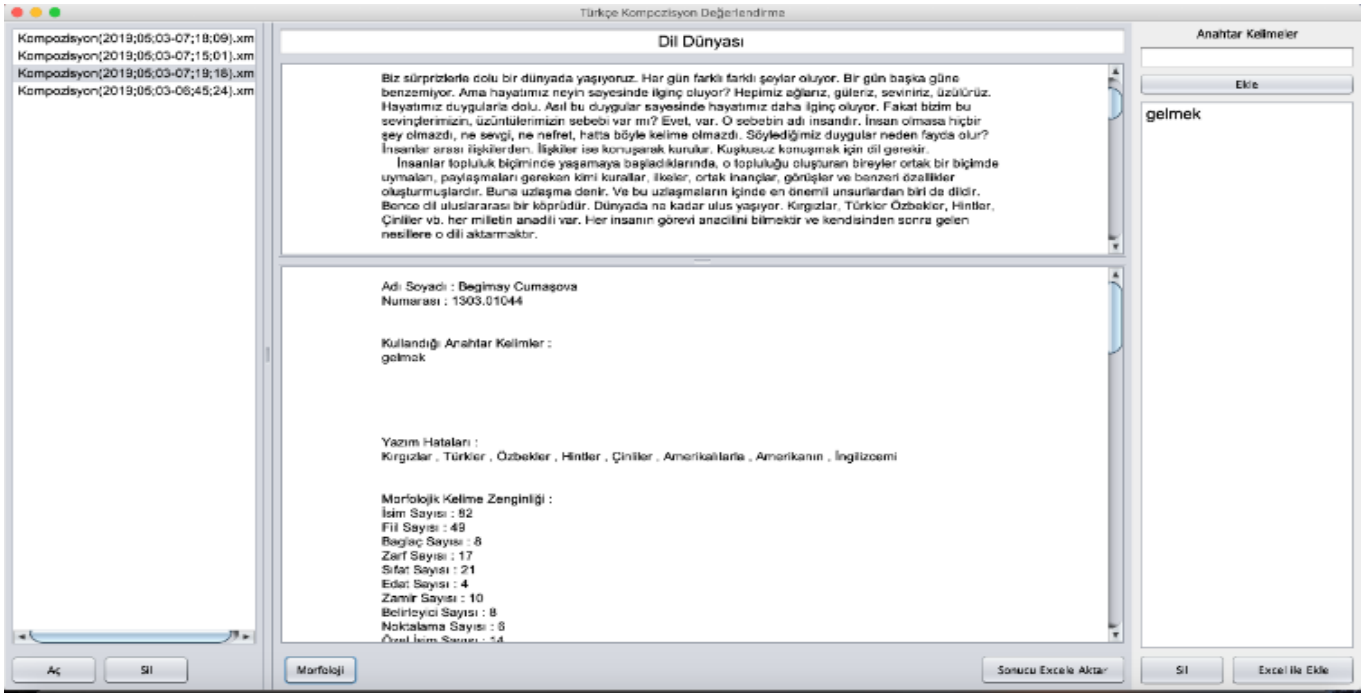

Figure 1. *The interface of the feature extraction tool for essay evaluation*

As can be seen from Table 1, an instructor can evaluate how much the student mastered the subject based on the use (or disuse) of keywords. In the output table, the use of keywords is indicated. The spelling errors are also listed.

The semantic structure of sentences is extracted using the Zemberek library, according to which, the count of parts of the speech is carried out. However, despite the normalization process, the proposed system was unable to determine the part of speech of some words. Yet, the success rate was approximately 98%, with maximum 2 words with undefined classification at each of tested essays.

In addition, comparing the total number of unique words that a student used in an essay with the total number of words used, an instructor can measure whether the student constantly makes statements using the same words. In addition, the number of sentences, where the modalities of indicative moods and subjunctive moods were used, allows an instructor to measure how many moods the student can use and the ability to use certain Turkish suffixes.

5. Conclusion

In this work, we present a simple tool that can help teachers to evaluate students' essays. Although there are studies in the field of Turkish natural language processing, there is no study such as the evaluation of an essay written by Turkish language learners with natural language processing methods. As mentioned before, at the current stage of natural language processing techniques, it is infeasible to implement some text evaluation measurements. Instead, as many works suggest, for further implementation of statistical machine learning techniques, it is necessary to extract text features. Therefore, in current work, in addition to assisting language instructors, the list of essay evaluation scales was proposed and software was developed to extract these features. Yet, the list of features was limited to six measurements.

The limitation of the current work is that some evaluation scales such as relevance or redundancy of words and sentences or disclosure of the topic of the essay were not implemented. The difficulty of this task was widely discussed in the literature. Yet, as mentioned above, in the scope of the current work, the attempt to solve this issue was done by counting usage of keywords, provided by language instructor for a given topic.

Nevertheless, these measurements can more or less predict the grade given to an essay. Therefore, as future work, it is planned to use the developed software and extracted text features for building a model for automatic essay evaluation using machine learning techniques.

References

[1]. Akın A.A., Akın M.D. “Zemberek, an open-source NLP framework for Turkic languages”, Structure, 10, (2007), 1-5.

[2]. Bahdanau D., Cho K., Bengio Y. “Neural machine translation by jointly learning to align and translate”, arXiv preprint arXiv:1409.0473, (2014).

[3]. Bird S., Klein E., Loper E. “Natural language processing with Python: analyzing text with the natural language toolkit”, Sebastopol: O'Reilly Media, Inc., (2009).

[4]. Brants T., Popat A.C., Xu P., Och F.J., Dean J. “Large language models in machine translation”, In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), (2007, June), 858-867.

[5]. Brown P.F., Desouza P.V., Mercer R.L., Pietra V.J. D., Lai J.C., “Class-based n-gram models of natural language”, Computational linguistics, 18(4), (1992), 467-479.

[6]. Burstein J., Chodorow M., Leacock C. “CriterionSM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays”, In IAAI, (2003, August), 3-10.

[7]. Cavnar W.B., Trenkle, J.M. “N-gram-based text categorization”, In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, Vol. 161175, (1994, April).

[8]. Chowdhury G.G. Natural language processing”, Annual review of information science and technology, 37(1), (2003), 51-89.

[9]. Collobert R., Weston J. “A unified architecture for natural language processing: Deep neural networks with multitask learning”, In Proceedings of the 25th international conference on Machine learning. ACM, (2008, July), 160-167.

[10]. Goyal A., Jagarlamudi J. Daumé III, H., & Venkatasubramanian, S. “Sketching techniques for large scale NLP”, In Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop, (2010, June), 17-25. Association for Computational Linguistics.

[11]. Habash N.Y. “Introduction to Arabic natural language processing”, Synthesis Lectures on Human Language Technologies, 3(1), (2010), 1-187.

[12]. Khreisat L. “A machine learning approach for Arabic text classification using N-gram frequency statistics”, Journal of Informetrics, 3(1), (2009), 72-77.

[13]. Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran

Ch., Zens R., Dyer Ch., Bojar O., Constantin A., Herbst E. “Moses: Open source toolkit for statistical machine translation”, In Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions, (2007, June), 177-180.

[14]. Page E.B. “The imminence of... grading essays by computer”, The Phi Delta Kappan, 47(5), (1966), 238-243.

[15]. Page E.B. “Grading essays by computer: Progress report”, In Proceedings of the Invitational Conference on Testing Problems, (1967).

[16]. Page E.B. “Computer grading of student prose, using modern concepts and software”, The Journal of experimental education, 62(2), (1994), 127-142.

[17]. Page E.B. Poggio, J. P., Keith, T. Z., “Computer analysis of student essays: Finding trait differences in the student profile”, Paper presented at the annual meeting of the American Educational Research Association, Chicago, (1997, March).

[18]. Papineni K., Roukos S., Ward, T., Zhu W.J., “BLEU: a method for automatic evaluation of machine translation“, In Proceedings of the 40th annual meeting on association for computational linguistics, (2002, July), 311-318. Association for Computational Linguistics.

[19]. Schraudner M. “The online teacher’s assistant: Using automated correction programs to supplement learning and lesson planning”, CELE Journal, 22, (2014), 128-140.

[20]. Shermis M.D., Koch C.M., Page E.B., Keith T.Z., Harrington S. “Trait ratings for automated essay grading“, Educational and Psychological Measurement, 62(1), (2002), 5-18.

[21]. Stenetorp P., Pyysalo S., Topić G., Ohta T., Ananiadou S., Tsujii J.I. “BRAT: a web-based tool for NLP-assisted text annotation”, In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, (2012, April), 102-107. Association for Computational Linguistics.

[22]. Taghipour K., Ng H.T. “A neural approach to automated essay scoring”, In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, (2016, November), 1882-1891.

[23]. Tayirova N., Tekerek M., Brimkulov U. “Statistical machine translation implementation and performance tests between Kyrgyz and Turkish Languages”, MANAS Journal of Engineering, 3 (2), (2015), 59-68.

[24]. Tyers F.M., Alperen M.S. “South-east European times: A parallel corpus of Balkan languages”, In Proceedings of the LREC Workshop on Exploitation of Multilingual Resources and Tools for Central and (South-) Eastern European Languages, (2010), 49-53.

[25]. Van Ewijk R. “Same work, lower grade? Student ethnicity and teachers’ subjective assessments”, Economics of Education Review, 30(5), (2011), 1045-1058.

[26]. Zupanc K., Bosnić Z. “Automated essay evaluation with semantic analysis”, Knowledge-Based Systems, 120, (2017), 118-132

.